Cluster Reverberation: a mechanism for robust short-term memory without synaptic learning

Samuel Johnson, J. Marro, and Joaquín J. Torres
Departamento de Electromagnetismo y Física de la Materia, and
Institute Carlos I for Theoretical and Computational Physics,
University of Granada, E-18071 Granada, Spain.

Abstract

Short-term memory cannot in general be explained the way long-term memory can – as a gradual modification of synaptic conductances – since it takes place too quickly. Theories based on some form of cellular bistability, however, do not seem to be able to account for the fact that noisy neurons can collectively store information in a robust manner. We show how a sufficiently clustered network of simple model neurons can be instantly induced into metastable states capable of retaining information for a short time. Cluster Reverberation, as we call it, could constitute a viable mechanism available to the brain for robust short-term memory with no need of synaptic learning. Relevant phenomena described by neurobiology and psychology, such as power-law statistics of forgetting avalanches, emerge naturally from this mechanism.

Keywords: Working memory, sensory memory, iconic memory, power-law forgetting, nonequilibrium neural networks.

1 Slow but sure, or fast and fleeting?

Of all brain phenomena, memory is probably one of the best understood [1, 2, 3]. Consider a set of many neurons, defined as elements with two possible states (firing or not firing, one or zero) connected among each other in some way by synapses which carry a proportion of the current let off by a firing neuron to its neighbours; the probability that a given neuron has of firing at a certain time is then some function of the total current it has just received. Such a simplified model of the brain is able to store and retrieve information, in the form of patterns of activity (i.e., particular configurations of firing and non-firing neurons) when the synaptic conductances, or weights, have been appropriately set according to a learning rule [4]. Because each of the stored patterns becomes an attractor of the dynamics, the system will evolve towards whichever of the patterns most resembles the initial configuration. Artificial systems used for tasks such as pattern recognition and classification, as well as more realistic neural network models that take into account a variety of subcellular processes, all tend to rely on this basic mechanism, known as Associative Memory [5, 6].

Synaptic conductances in animal brains have indeed been found to become strengthened or weakened during learning, via the biochemical processes of long-

term potentiation (LTP) and depression (LTD) [7, 8]. Further support for the hypothesis that such a mechanism underlies long-term memory (LTM) comes from psychology, where it is being found more and more that so-called *connectionist* models fit in well with observed brain phenomena [9, 10]. However, some memory processes take place on timescales of seconds or less and in many instances cannot be accounted for by LTP and LTD [11], since these require at least minutes to be effected [12, 13]. For example, Sperling found that visual stimuli are recalled in great detail for up to about one second after exposure (iconic memory) [14]; similarly, acoustic information seems to linger for three or four seconds (echoic memory) [15]. In fact, it appears that the brain actually holds and continually updates a kind of buffer in which sensory information regarding its surroundings is maintained (sensory memory) [16]. This is easily observed by simply closing one's eyes and recalling what was last seen, or thinking about a sound after it has finished. Another instance is the capability referred to as *working* memory [11, 17]: just as a computer requires RAM for its calculations despite having a hard drive for long term storage, the brain must continually store and delete information to perform almost any cognitive task. To some extent, working memory could consist in somehow labelling or bringing forward previously stored concepts, like when one is asked to remember a particular sequence of digits or familiar shapes. But we are also able to manipulate – if perhaps not quite so well – shapes and symbols we have only just become acquainted with, too recently for them to have been learned synaptically. We shall here use *short-term* memory (STM) to describe the brain's ability to store information on a timescale of seconds or less[1].

Evidence that short-term memory is related to sensory information while long-term memory is more conceptual can again be found in psychology. For instance, a sequence of similar sounding letters is more difficult to retain for a short time than one of phonetically distinct ones, while this has no bearing on long-term memory, for which semantics seems to play the main role [18, 19]; and the way many of us think about certain concepts, such as chess, geometry or music, is apparently quite sensorial: we imagine positions, surfaces or notes as they would look or sound. Most theories of short-term memory – which almost always focus on working memory – make use of some form of previously stored information (i.e., of synaptic learning) and so can account for the labelling tasks referred to above but not for the instant recall of novel information [20, 21, 22, 23]. Attempts to deal with the latter have been made by proposing mechanisms of *cellular bistability*: neurons are assumed to retain the state they are placed in (such as firing or not firing) for some period of time thereafter [24, 25, 26]. Although there may indeed be subcellular processes leading to a certain bistability, the main problem with short-term memory depending exclusively on such a mechanism is that if each neuron must act independently of the rest the patterns will not be robust to random fluctuations [11] – and the behaviour of individual neurons is known to be quite noisy [27]. It is worth pointing out that

[1] We should mention that sensory memory is usually considered distinct from STM – and probably has a different origin – but we shall use "short-term memory" generically since the mechanism we propose in this paper could be relevant for either or both phenomena. On the other hand, the recent flurry of research in psychology and neuroscience on working memory has lead to this term sometimes being used to mean short-term memory; strictly speaking, however, working memory is generally considered to be an aspect of cognition which operates on information stored in STM.

one of the strengths of Associative Memory is that the behaviour of a given neuron depends on many neighbours and not just on itself, which means that robust global recall can emerge despite random fluctuations at an individual level.

Something that, at least until recently, most neural network models have failed to take into account is the structure of the network – its topology – it often being assumed that synapses are placed among the neurons completely at random, or even that all neurons are connected to all the rest (a mathematically convenient but unrealistic situation). Although relatively little is yet known about the architecture of the brain at the level of neurons and synapses, experiments have shown that it is heterogeneous (some neurons have very many more synapses than others), clustered (two neurons have a higher chance of being connected if they share neighbours than if not) and highly modular (there are groups, or modules, with neurons forming synapses preferentially to those in the same module) [28, 29]. We show here that it suffices to use a more realistic topology, in particular one which is modular and/or clustered, for a randomly chosen pattern of activity the system is placed in to be metastable. This means that novel information can be instantly stored and retained for a short period of time in the absence of both synaptic learning and cellular bistability. The only requisite is that the patterns be coarse grained versions of the usual patterns – that is, whereas it is often assumed that each neuron in some way represents one bit of information, we shall allocate a bit to a small group or neurons[2] (four or five can be enough).

The mechanism, which we call Cluster Reverberation, is very simple. If neurons in a group are more highly connected to each other than to the rest of the network, either because they form a module or because the network is significantly clustered, they will tend to retain the activity of the group: when they are all initially firing, they each continue to receive many action potentials and so go on firing, whereas if they start off silent, there is not usually enough input current from the outside to set them off. The fact that each neuron's state depends on its neighbours conferres to the mechanism a certain robustness in the face of random fluctuations. This robustness is particularly important for biological neurons, which as mentioned are quite noisy. Furthermore, not only does the limited duration of short-term memory states emerge naturally from this mechanism (even in the absence of interference from new stimuli) but this natural forgetting follows power-law statistics, as in experimental settings [30, 31, 32].

The process is reminiscent both of block attractors in ordinary neural networks [33] and of domains in magnetic materials [34], while Muñoz et al. have recently highlighted a similarity with Griffiths phases on networks [35]. It can also be interpreted as a multiscale phenomenon: the mesoscopic clusters take on the role usually played by individual neurons, yet make use of network properties. Although the mechanism could also work in conjunction with other ones, such as synaptic learning or cellular bistability, we shall illustrate it by considering the simplest model which has the necessary ingredients: a set of binary neurons linked by synapses of uniform weight according to a topology whose modularity or clustering we shall tune. As with Associative Memory, this mech-

[2]This does not, of course, mean that memories are expected to be encoded as bitmaps. Just as with individual neurons, positions or orientations, say, could be represented by the activation of particular sets of clusters.

anism of Cluster Reverberation appears to be simple and robust enough not to be qualitatively affected by the complex subcellular processes incorporated into more realistic neuron models – such as integrate-and-fire or Hodgkin-Huxley neurons. However, such refinements are probably needed to achieve graded persistent activity, since the mean frequency of each cluster could then be set to a certain value.

2 The simplest neurons on modular networks

We consider a network of N model neurons, with activities $s_i = \pm 1$. The topology is given by the adjacency matrix $\hat{a}_{ij} = \{1, 0\}$, each element representing the existence or absence of a synapse from neuron j to neuron i ($\hat{a}$ need not be symmetric). In this kind of model, each edge usually has a *synaptic weight* associated, $\omega_{ij} \in \mathbb{R}$; however, we shall here consider these to have all the same value: $\omega_{ij} = \omega \ \forall i, j$. Neurons are updated in parallel (Little dynamics) at each time step, according to the stochastic transition rule

$$P(s_i \to \pm 1) = \pm \frac{1}{2} \tanh\left(\frac{h_i}{T}\right) + \frac{1}{2},$$

where the *field* of neuron i is defined as

$$h_i = \omega \sum_j^N \hat{a}_{ij} s_j$$

and T is a parameter we shall call *temperature*.

First of all, we shall consider the network defined by $\hat{a}$ to be made up of M distinct modules. To achieve this, we can first construct M separate random directed networks, each with $n = N/M$ nodes and mean degree (mean number of neighbours) $\langle k \rangle$. Then we evaluate each edge and, with probability λ, eliminate it, to be substituted for another edge between the original post-synaptic neuron and a new pre-synaptic neuron chosen at random from among any of those in other modules[3]. Note that this protocol does not alter the number of pre-synaptic neighbours of each node, $k_i^{in} = \sum_j \hat{a}_{ij}$ (although the number of post-synaptic neurons, $k_i^{out} = \sum_j \hat{a}_{ji}$, can vary). The parameter λ can be seen as a measure of *modularity* of the partition considered, since it coincides with the expected value of the proportion of edges that link different modules. In particular, $\lambda = 0$ defines a network of disconnected modules, while $\lambda = 1 - M^{-1}$ yields a random network in which this partition has no modularity. If $\lambda \in (1 - M^{-1}, 1)$, the partition is less than randomly modular – i.e., it is *quasi-multipartite* (or multipartite if $\lambda = 1$).

If the size of the modules is of the order of $\langle k \rangle$, the network will also be highly clustered. Taking into account that the network is directed, let us define the clustering coefficient C_i as the probability, given that there is a synapse from neuron i to a neuron j and from another neuron l to i, that there be a synapse from j to l: that is, that there exist a feedback loop $i \to j \to l \to i$. Then,

[3] We do not allow self-edges (although these can occur in reality) since they can be regarded as a form of cellular bistability.

assuming $M \gg 1$, the expected value of the clustering coefficient $C \equiv \langle C_i \rangle$ is

$$C \gtrsim \frac{\langle k \rangle - 1}{n-1}(1-\lambda)^3.$$

3 Cluster Reverberation

A memory pattern, in the form of a given configuration of activities, $\{\xi_i = \pm 1\}$, can be stored in this system with no need of prior learning. Imagine a pattern such that the activities of all n neurons found in any module are the same, i.e., $\xi_i = \xi_{\mu(i)}$, where the index $\mu(i)$ denotes the module that neuron i belongs to. This can be thought of as a coarse graining of the standard idea of memory patterns, in which each neuron represents one bit of information. In our scheme, each module represents – and stores – one bit. The system can be induced into this configuration via the application of an appropriate *stimulus* (see Fig. 1): the field of each neuron will be altered for just one time step according to

$$h_i \to h_i + \delta \xi_{\mu(i)}, \quad \forall i,$$

where the factor δ is the intensity of the stimulus. This mechanism for dynamically storing information will work for values of parameters such that the system is sensitive to the stimulus, acquiring the desired configuration, yet also able to retain it for some interval of time thereafter.

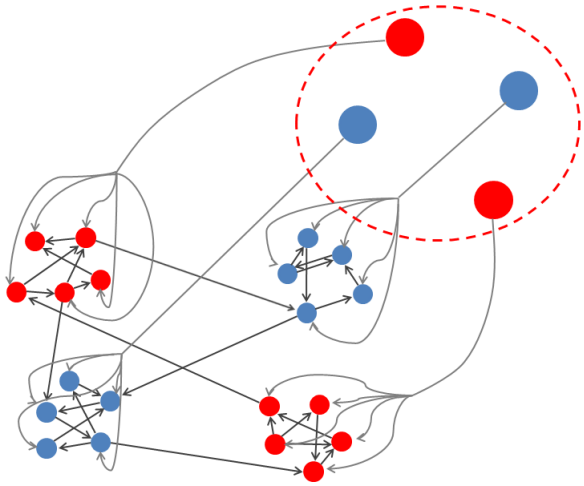

Figure 1: Diagram of a modular network composed of four five-neuron clusters. The four circles enclosed by the dashed line represent the stimulus: each is connected to a particular module, which adopts the input state (red or blue) and retains it after the stimulus has disappeared via Cluster Reverberation.

The two main attractors of the system are $s_i = 1 \ \forall i$ and $s_i = -1 \ \forall i$. These are the configurations of minimum energy (see the next section for a more detailed discussion on energy). However, the energy is locally minimised for any configuration in which $s_i = d_{\mu(i)} \ \forall i$ with $d_\mu = \pm 1$; that is, configurations

such that each module comprises either all active or all inactive neurons. These are the configurations that we shall use to store information. We define the mean activity[4] of each module,

$$m_\mu \equiv \frac{1}{n}\sum_{i\in\mu}^{n} s_i,$$

which is a mesoscopic variable, as well as the global mean activity,

$$m \equiv \frac{1}{N}\sum_{i}^{N} s_i = \frac{1}{M}\sum_{\mu}^{M} m_\mu$$

(these magnitudes change with time, but, where possible, we shall avoid writing the time dependence explicitely for clarity). The extent to which the network, at a given time, retains the pattern $\{\xi_i\}$ with which it was stimulated is measured with the *overlap* parameter

$$m_{stim} \equiv \frac{1}{N}\sum_{i}^{N} \xi_i s_i = \frac{1}{M}\sum_{\mu}^{M} \xi_\mu m_\mu.$$

Ideally, the system should be capable of reacting immediately to a stimulus by adopting the right configuration, yet also be able to retain it for long enough to use the information once the stimulus has disappeared. A measure of performance for such a task is therefore

$$\eta \equiv \frac{1}{\tau}\sum_{t=t_0+1}^{t_0+\tau} m_{stim}(t),$$

where t_0 is the time at which the stimulus is received and τ is the period of time we are interested in ($|\eta| \leq 1$) [38]. If the intensity of the stimulus, δ, is very large, then the system will always adopt the right pattern perfectly and η will only depend on how well it can then retain it. In this case, the best network will be one that is made up of unconnected modules. However, since the stimulus in a real brain can be expected to arrive via a relatively small number of axons, either from another part of the brain or directly from sensory cells, it is more realistic to assume that δ is of a similar order as the input a typical neuron receives from its neighbours, $\langle h\rangle \sim \omega\langle k\rangle$.

Fig. 2 shows the mean performance obtained when the network is repeatedly stimulated with different randomly generated patterns. For low enough values of the modularity λ and stimuli of intensity $\delta \gtrsim \omega\langle k\rangle$, the system can capture and successfully retain any pattern it is "shown" for some period of time, even though this pattern was in no way previously learned. For less intense stimuli ($\delta < \omega\langle k\rangle$), performance is nonmonotonic with modularity: there exists an optimal value of λ at which the system is sensitive to stimuli yet still able to retain new patterns quite well.

It is worth noting that performance can also break down due to thermal fluctuations. The two main attractors of the system ($s_i = 1\ \forall i$ and $s_i = -1$

[4]The mean activity in a neural network model is usually taken to represent the mean firing rate measured in experiments [3].

$\forall i$) suffer the typical second order phase transition of the Hopfield model [5], from a memory phase (one in which $m = 0$ is not stable and stable solutions $m \neq 0$ exist) to one with no memory (with $m = 0$ the only stable solution), at the critical temperature [38]

$$T_c = \omega \frac{\langle k_{in}^2 \rangle}{\langle k \rangle}.$$

(Note that, in a directed network, $\langle k_{in} \rangle = \langle k_{out} \rangle \equiv \langle k \rangle$, although the other moments can in general be different.) The metastable states we are interested in, though, have a critical temperature

$$T_c' = (1 - \lambda) T_c$$

(assuming that the mean activity of the network is $m \simeq 0$). That is, the temperature at which the modules are no longer able to retain their individual activity is in general lower than that at which the the solution $m = 0$ for the whole network becomes stable.

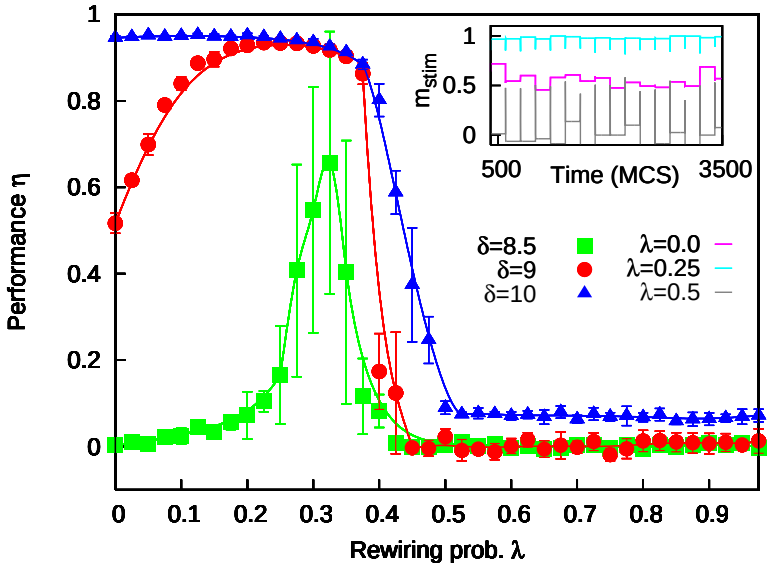

Figure 2: Performance η against λ for networks of the sort described in the main text with $M = 160$ modules of $n = 10$ neurons, $\langle k \rangle = 9$; patterns are shown with intensities $\delta = 8.5$, 9 and 10, and $T = 0.02$ (lines – splines – are drawn as a guide to the eye). Inset: typical time series of m_{stim} (i.e., the overlap with whichever pattern was last shown) for $\lambda = 0.0$, 0.25, and 0.5, and $\delta = \langle k \rangle = 9$.

4 Energy and topology

Each pair of nodes contributes a configurational energy $e_{ij} = -\omega \frac{1}{2}(\hat{a}_{ij} + \hat{a}_{ji}) s_i s_j$; that is, if there is an edge from i to j and they have opposite activities, the energy is increased in $\frac{1}{2}\omega$, whereas it is decreased by the same amount if their activities

are the same. Given a configuration, we can obtain its associated energy by summing over all pairs. We shall be interested in configurations with x neurons that have $s = +1$ (and $N - x$ with $s = -1$), chosen in such a way that one module at most, say μ, has neurons in both states simultaneously. Therefore, $x = n\rho + z$, where ρ is the number of modules with all their neurons in the positive state and z is the number of neurons with positive sign in module μ. We can write $m = (2x - 1)/N$ and $m_\mu = (2z - 1)/n$. The total configurational energy of the system will be $E = \sum_{ij} e_{ij} = \frac{1}{2}\omega(L_{\uparrow\downarrow} - \langle k \rangle N)$, where $L_{\uparrow\downarrow}$ is the number of edges linking nodes with opposite activities. By simply counting over edges, we can obtain the expected value of $L_{\uparrow\downarrow}$ (which amounts to a mean-field approximation because we are substituting the number of edges between two neurons for its expected value), yielding:

$$\frac{E}{\omega\langle k \rangle} = (1 - \lambda)\frac{z(n - z)}{n - 1}$$
$$+\frac{\lambda n}{N - n}\{\rho[n - z + n(M - \rho - 1)] + (M - \rho - 1)(z + n\rho)\} - \frac{1}{2}N. \qquad (1)$$

Fig. 3 shows the mean-field configurational energy curves for various values of the modularity on a small modular network. The local minima (metastable states) are the configurations used to store patterns. It should be noted that the mapping $x \to m$ is highly degenerate: there are C^M_{mM} patterns with mean activity m that all have the same energy.

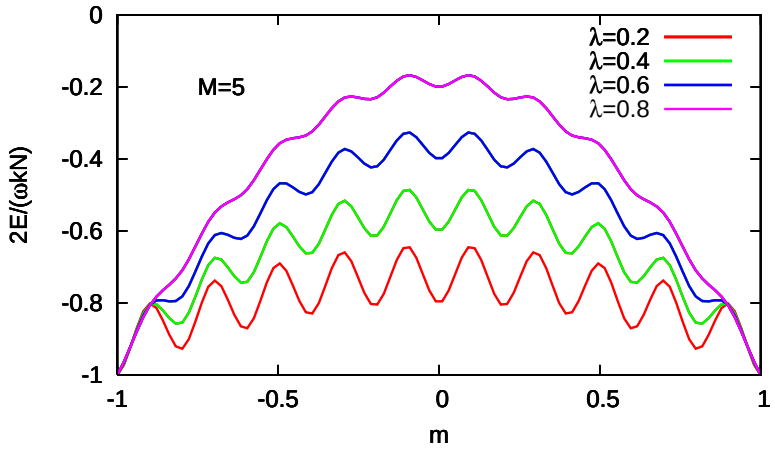

Figure 3: Configurational energy of a network composed of $M = 20$ modules of $n = 10$ neurons each, according to Eq. (1), for various values of the rewiring probability λ. The minima correspond to situations such that all neurons within any given module have the same sign.

5 Forgetting avalanches

In obtaining the energy we have assumed that the number of synapses rewired from a given module is always $\nu = \langle k \rangle n\lambda$. However, since each edge is evaluated

with probability λ, ν will in fact vary somewhat from one module to another, being approximately Poisson distributed with mean $\langle\nu\rangle = \langle k\rangle n\lambda$. The depth of the energy well corresponding to a given module is then, neglecting all but the first term in Eq. (1) and approximating $n - 1 \simeq n$,

$$\Delta E \simeq \frac{1}{4}\omega(n\langle k\rangle - \nu).$$

The typical escape time τ from an energy well of depth ΔE at temperature T is $\tau \sim e^{\Delta E/T}$ [36]. Using Stirling's approximation in the Poissonian distribution of ν and expressing it in terms of τ, we find that the escape times are distributed according to

$$P(\tau) \sim \left(1 - \frac{4T}{\omega n\langle k\rangle}\ln\tau\right)^{-\frac{3}{2}} \tau^{-\beta(\tau)}, \tag{2}$$

where

$$\beta(\tau) = 1 + \frac{4T}{\omega n\langle k\rangle}\left[1 + \ln\left(\frac{\lambda n\langle k\rangle}{1 - \frac{4T}{\omega n\langle k\rangle}\ln\tau}\right)\right]. \tag{3}$$

Therefore, at low temperatures, $P(\tau)$ will behave approximately like a power-law. The left panel of Fig. 4 shows the distribution of time intervals between events in which the overlap m_μ of at least one module μ changes sign. The power-law-like behaviour is apparent, and justifies talking about *forgetting avalanches* – since there are cascades of many forgetting events interspersed with long periods of metastability. This is very similar to the behaviour observed in other nonequilibrium settings in which power-law statistics arise from the convolution of exponentials [37, 35].

It is known from experimental psychology that forgetting in humans is indeed well described by power-laws [30, 31, 32]. The right panel of Fig. 4 shows the value of the exponent $\beta(\tau)$ as a function of τ. Although for low temperatures it is almost constant over many decades of τ – approximating a pure power-law – for any finite T there will always be a τ such that the denominator in the logarithm of Eq. (3) approaches zero and β diverges, signifying a truncation of the distribution.

6 Clustered networks

Although we have illustrated how the mechanism of Cluster Reverberation works on a modular network, it is not actually necessary for the topology to have this characteristic – only for the patterns to be in some way "coarse-grained," as described, and that each region of the network encoding one bit have a small enough parameter λ, defined as the proportion of synapses to other regions. For instance, for the famous Watts-Strogatz *small-world* model [39] – a ring of N nodes, each initially connected to its k nearest neighbours before a proportion p of the edges are randomly rewired – we have $\lambda \simeq p$ (which is not surprising considering the resemblance between this model and the modular network used above). More precisely, the expected modularity of a randomly imposed box of n neurons is

$$\lambda = p - \frac{n-1}{N-1}p + \frac{1-p}{n}\left(\frac{k}{4} - \frac{1}{2}\right),$$

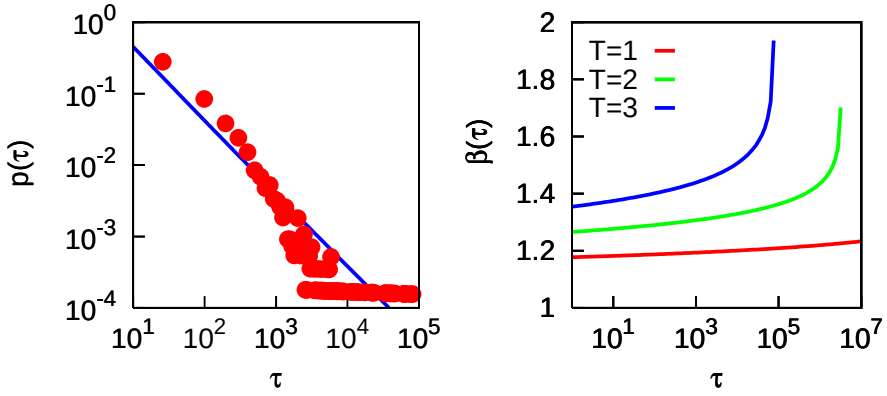

Figure 4: Left panel: distribution of escape times τ, as defined in the main text, for $\lambda = 0.22$ and $T = 0.02$. Other parameters as in Fig. 2. Symbols from MC simulations and line given by Eqs. (2) and (3). Right panel: exponent β of the quasi-power-law distribution $p(\tau)$ as given by Eq. (3) for temperatures $T = 1$ (red line), $T = 2$ (green line) and $T = 3$ (blue line).

the second term on the right accounting for the edges rewired to the same box, and the third to the edges not rewired but sufficiently close to the border to connect with a different box.

Perhaps a more realistic model of clustered network would be a random network embedded in d-dimensional Euclidean space. For this we shall use the scheme laid out by Rozenfeld *et al.* [40], which consists simply in allocating each node to a site on a d-torus and then, given a particular degree sequence, placing edges to the nearest nodes possible – thereby attempting to minimise total edge length[5]. For a scale-free degree sequence (i.e., a set $\{k_i\}$ drawn from a degree distribution $p(k) \sim k^{-\gamma}$) according to some exponent γ, then, as shown in Appendix 1, such a network has a modularity

$$\lambda \simeq \frac{1}{d(\gamma-2)-1}\left[d(\gamma-2)l^{-1} - l^{-d(\gamma-2)}\right], \tag{4}$$

where l is the linear size of the boxes considered.

Fig. 5 compares this expression with the value obtained numerically after averaging over many network realizations, and shows that it is fairly good – considering the approximations used for its derivation. It is interesting that even in this scenario, where the boxes of neurons which are to receive the same stimulus are chosen at random with no consideration for the underlying topology, these boxes need not have very many neurons for λ to be quite low (as long as the degree distribution is not too heterogeneous).

Carrying out the same repeated stimulation test as on the modular networks in Fig. 2, we find a similar behaviour for the scale-free embedded networks. This is shown in Fig. 6, where for high enough intensity of stimuli δ and scale-free exponent γ, performance can, as in the modular case, be $\eta \simeq 1$. We should point out that for good performance on these networks we require more neurons for

[5]The authors also consider a cutoff distance, but we shall take this to be infinite here.

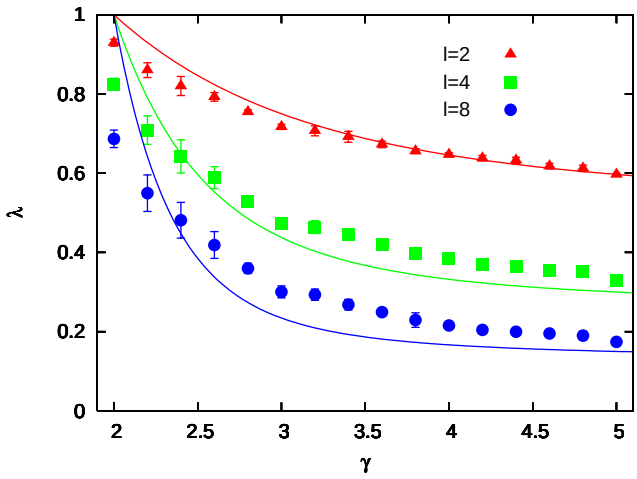

Figure 5: Proportion of outgoing edges, λ, from boxes of linear size l against exponent γ for scale-free networks embedded on $2D$ lattices. Lines from Eq. (4) and symbols from simulations with $\langle k \rangle = 4$ and $N = 1600$.

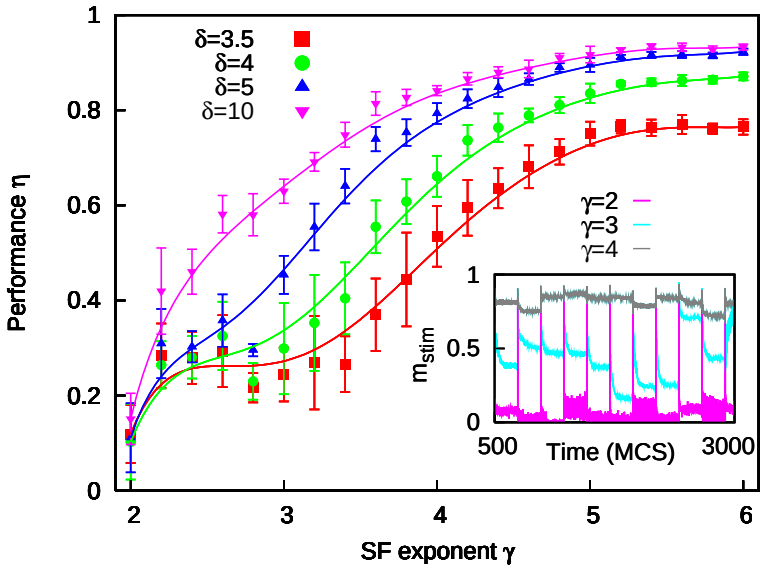

Figure 6: Performance η against exponent γ for scale-free networks, embedded on a 2D lattice, with patterns of $M = 16$ modules of $n = 100$ neurons each, $\langle k \rangle = 4$ and $N = 1600$; patterns are shown with intensities $\delta = 3.5$, 4, 5 and 10, and $T = 0.01$ (lines – splines – are drawn as a guide to the eye). Inset: typical time series for $\gamma = 2$, 3, and 4, with $\delta = 5$.

each bit of information than on modular networks with the same λ (in Fig. 6 we use $n = 100$, as opposed to $n = 10$ in Fig. 2). However, that we should be able

to obtain good results for such diverse network topologies underlines that the mechanism of Cluster Reverberation is robust and not dependent on some very specific architecture. In fact, we have recently shown that similar metastable memory states can also occur on networks which have random modularity and clustering, but a certain degree of *assortativity*[6] [42].

7 Yes, but does it happen in the brain?

As we have shown, Cluster Reverberation is a mechanism available to neural systems for robust short-term memory without synaptic learning. To the best of our knowledge, this is the first mechanism proposed which has these characteristics – essential for, say, sensory memory or certain working-memory tasks. All that is needed is for the network topology to be highly clustered or modular, and for small groups of neurons to store one bit of information, as opposed to the conventional view which assumes one bit per neuron. Considering the enormous number of neurons in the brain, and the fact that real individual neurons are probably too noisy to store information reliably, these hypotheses do not seem farfetched. The mechanism is furthermore consistent both with what is known about the topology of the brain, and with experiments which have revealed power-law forgetting.

Since the purpose of this paper is only to describe the mechanism of Cluster Reverberation, we have made use of the simplest possible model neurons – namely, binary neurons with static, uniform synapses – for the sake of clarity and generality. However, there is no reason to believe that the mechanism would cease to function if more neuronal ingredients were to be incorporated. In fact, cellular bistability, for instance, would increase performance, and the two mechanisms could actually work in conjunction. Similarly, we have also used binary patterns to store information. But it is to be expected that patterns depending on any form of frequency coding, for instance, could also be maintained with more sophisticated neurons – such that different modules could be set to different mean frequencies.

Whether Cluster Reverberation would work for biological neural systems could be put to the test by growing such modular networks *in vitro*, stimulating appropriately, and observing the duration of the metastable states. *In vivo* recordings of neural activity during short-term memory tasks, together with a mapping of the underlying synaptic connections, could be used to ascertain whether the brain does indeed make use of this mechanism – although for this it must be borne in mind that the neurons forming a module need not find themselves close together in metric space. We hope that experiments such as these will be carried out and eventually reveal something more about the basis of this puzzling emergent property of the brain's known as thought.

Acknowledgements

This work was supported by Junta de Andalucía projects FQM-01505 and P09-FQM4682, and by Spanish MEC-FEDER project FIS2009-08451. Many thanks

[6] The assortativity of a network is here understood to mean the extent to which the degrees of neighbouring nodes are correlated [41].

to Jorge F. Mejias, Sebastiano de Franciscis and Miguel A. Muñoz for stimulating and didactic conversations.

References

[1] Amit, D.J., *Modeling Brain Function*, Cambridge Univ. Press, Cambridge, 1989.

[2] L.F. Abbott and T.B. Kepler, "From Hodgkin-Huxley to Hopfield", *Statistical mechanics of neural networks*, Springer-Verlag, Berlin, 1990.

[3] J.J. Torres, and P. Varona, "Modeling Biological Neural Networks", *Handbook of Natural Computing*, Springer-Verlag, Berlin, 2010.

[4] Hebb, D.O., *The organization of behavior*, Wiley, New York, 1949.

[5] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", *Proc. Natl. Acad. Sci. USA* **79** 2554–8 (1982).

[6] S. Amari, "Characteristics of random nets of analog neuron-like elements", *IEEE Trans. Syst. Man. Cybern.,* **2** 643–657 (1972).

[7] A. Gruart, M.D. Muñoz, and J.M. Delgado-García, "Involvement of the CA3-CA1 synapse in the acquisition of associative learning in behaving mice", *J. Neurosci.* **26**, 1077–87 (2006).

[8] M. De Roo, P. Klauser, P. Mendez, L. Poglia, and D. Muller, "Activity-dependent PSD formation and stabilization of newly formed spines in hippocampal slice cultures", *Cerebral Cortex* **18**, 151 (2008).

[9] Marcus, G.F., *The Algebraic Mind: Integrating Connectionism and Cognitive Science*, MIT Press, Cambrigde, MA, 2001.

[10] S.L. Frank, W.F.G. Haselager, and I. van Rooij, "Connectionist semantic systematicity", *Cognition* **110**, 358-79 (2009).

[11] D. Durstewitz, J.K. Seamans, and T.J. Sejnowski, "Neurocomputational models of working memory", *Nature Neuroscience* **3** 1184–91 (2000).

[12] K.S. Lee, F. Schottler, M. Oliver, and G. Lynch, "Brief bursts of high-frequency stimulation produce two types of structural change in rat hippocampus", *J. Neurophysiol.* **44**, 247 (1980).

[13] A.Y. Klintsova and W.T. Greenough, "Synaptic plasticity in cortical systems", *Current Opinion in Neurobiology* **9**, 203 (1999).

[14] G.A. Sperling, "The information available in brief visual persentation", *Psychological Monographs: General and Applied* **74**, 1–30 (1960).

[15] N. Cowan, "On Short And Long Auditory Stores", *Psychological Bulletin* **96**, 341–70 (1984).

[16] Baddeley, A. D., *Essentials of Human Memory*, Psychology Press, 1999.

[17] A. Baddeley, "Working memory: looking back and looking forward", *Nature Reviews Neuroscience* **4**, 829-39 (2003).

[18] R. Conrad, "Acoustic confusion in immediate memory", *B. J. Psychol.* **55**, 75–84 (1964).

[19] R. Conrad, "Information, acoustic confusion and memory span", *B. J. Psychol.* **55**, 429–432 (1964).

[20] X.-J. Wang, "Synaptic reverbaration underlying mnemonic presistent activity", *TRENDS in Neurosci.* **24**, 455–63 (2001).

[21] O. Barak and M. Tsodyks, "Persistent activity in neural networks with dynamic synapses", *PLoS Comput. Biol.* **3**(2): e35.

[22] Y. Roudi and P.E. Latham, "A balanced memory network", *PLoS Comput. Biol.* **3**(9): e141 (2007).

[23] G. Mongillo, O. Barak, and M. Tsodyks, "Synaptic Theory of Working Memory", *Science* **319**, 1543–1546 (2008).

[24] M. Camperi and X.-J. Wang, "A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability", *J. Comp. Neurosci.* **5**, 383–405 (1998).

[25] J.-N. Teramae and T. Fukai, "A cellular mechanism for graded persistent activity in a model neuron and its implications for working memory", *J. Comput. Neurosci.* **18**, 105–21 (2005).

[26] E. Tarnow, "Short Term Memory May Be the Depletion of the Readily Releasable Pool of Presynaptic Neurotransmitter Vesicles", *Cognitive Neurodynamics* (2008).

[27] A. Compte, C. Constantinidis, J. Tegner, S. Raghavachari, M.V. Chafee, *et al.* "Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task", *J. Neurophysiol.* **90** 3441-54 (2003).

[28] O. Sporns, D.R. Chialvo, M. Kaiser and C.C. Hilgetag, "Organization, development and function of complex brain networks", *Trends Cogn. Sci.* **8** 418–25 (2004).

[29] S. Johnson, J. Marro, and J.J. Torres, "Evolving networks and the development of neural systems", *J. Stat. Mech.* (2010) P03003.

[30] J.T. Wixted and E.B. Ebbesen, "On the form of forgetting", *Psychological Science* **2** 409–15 (1991).

[31] J.T. Wixted and E.B. Ebbesen, "Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions", *Memory & Cognition* **25**, 731–9 (1997).

[32] S. Sikström, "Forgetting curves: implications for connectionist models", *Cognitive Psychology* **45**, 95–152 (2002).

[33] D. Dominguez, M. González, E Serrano, and F.B. Rodríguez, "Structured information in small-world neural networks", *Phys. Rev. E* **79**, 021909 (2009).

[34] Hubert A. and Shaefer R. *Magnetic Domains*, Springer, Berlin, 1998.

[35] M.A. Muñoz, R. Juhasz, C. Castellano, G. Odor, "Griffiths phases in complex networks", preprint (2010).

[36] Levine, R.D., *Molecular Reaction Dynamics*, Cambridge University Press, Cambridge, 2005.

[37] P.I. Hurtado, J. Marro, and P.L. Garrido, "Demagnetization via nucleation of the nonequilibrium metastable phase in a model of disorder", *J. Stat. Phys.* **133**, 29-58 (2008)

[38] S. Johnson, J. Marro, and J.J. Torres, "Functional optimization in complex excitable networks", *EPL* **83**, 46006 (2008).

[39] D.J. Watts and S.H. Strogatz, "Collective dynamics of 'small-world' networks" *Nature* **395**, 440–2 (1998).

[40] A.F. Rozenfeld, R. Cohen, D. ben-Avraham, and S. Havlin, "Scale-free networks on lattices", *Phys. Rev. Lett.* **89**, 218701 (2002).

[41] S. Johnson, J.J. Torres, J. Marro, and M.A. Muñoz, "Entropic origin of disassortativity in complex networks", *Phys. Rev. Lett.* **104**, 108702 (2010).

[42] S. de Franciscis, S. Johnson, and J.J. Torres, "Influence of assortativity on attractor-neural-network performance", *in preparation.*

[43] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", *Science* **286**, 509–12 (1999).

Appendix 1

The number of nodes within a radius r is $n(r) = A_d r^d$, with A_d a constant. We shall therefore assume a node with degree k to have edges to all nodes up to a distance $r(k) = (k/A_d)^{1/d}$, and none beyond (note that this is not necessarily always feasible in practice). To estimate λ, we shall first calculate the probability that a randomly chosen edge have length x. The chance that the edge belong to a node with degree k is $\pi(k) \sim kp(k)$ (where $p(k)$ is the degree distribution). The proportion of edges that have length x among those belonging to a node with degree k is $\nu(x|k) = dA_d x^{d-1}/k$ if $A_d x^d < k$, and 0 otherwise. Considering, for example, scale-free networks (as in Ref. [40]), so that the degree distribution is $p(k) \sim k^{-\gamma}$ in some interval $k \in [k_0, k_{max}]$ [43], and integrating over $p(k)$, we have the distribution of lengths,

$$P(x) = (Const.) \int_{\max(k_0, Ax^d)}^{k_{max}} \pi(k)\nu(k|x)dk = d(\gamma - 2)x^{-[d(\gamma-2)+1]},$$

where we have assumed, for simplicity, that the network is sufficiently sparse that $\max(k_0, Ax^d) = Ax^d$, $\forall x \geq 1$, and where we have normalised for the

interval $1 \leq x < \infty$; strictly, $x \leq (k_{max}/A)^{1/d}$, but we shall also ignore this effect. Next we need the probability that an edge of length x fall between two compartments of linear size l. This depends on the geometry of the situation as well as dimensionality; however, a first approximation which is independent of such considerations is

$$P_{out}(x) = \min\left(1, \frac{x}{l}\right).$$

We can now estimate the modularity λ as

$$\lambda = \int_1^\infty P_{out}(x)P(x)dx = \frac{1}{d(\gamma-2)-1}\left[d(\gamma-2)l^{-1} - l^{-d(\gamma-2)}\right].$$

Fig. 5 shows how λ depends on γ for $d = 2$ and various box sizes.